University of Michigan School of Public Health

The University of Michigan Department of Biostatistics Working Paper Series

Year 2004 *Paper* 24

Individual Prediction in Prostate Cancer Studies Using a Joint Longitudinal-Survival-Cure Model

Menggang Yu* Jeremy Taylor†

Howard M. Sandler‡

*University of Michigan Biostatistics

†University of Michigan, jmgt@umich.edu

‡University of Michigan Radiation Oncology

http://biostats.bepress.com/umichbiostat/paper24

Individual Prediction in Prostate Cancer Studies Using a Joint Longitudinal-Survival-Cure Model

Menggang Yu, Jeremy Taylor, and Howard M. Sandler

Abstract

For monitoring patients treated for prostate cancer, Prostate Specific Antigen (PSA) is measured periodically after they receive treatment. Increases in PSA are suggestive of recurrence of the cancer and are used in making decisions about possible new treatments. The data from studies of such patients typically consist of longitudinal PSA measurements, censored event times and baseline covariates. Methods for the combined analysis of both longitudinal and survival data have been developed in recent years, with the main emphasis being on modeling and estimation. We analyze data from a prostate cancer study that has been extended by adding a mixture structure to the survival model component of the model. Here we focus on utilizing the model to make individualized prediction of disease progression for censored and alive patients.

In this model each patient is assumed to be either cured by the treatment or susceptible to clinical recurrence. The cured fraction is modeled as a logistic function of baseline covariates, measured before the end of the radiation therapy period. The longitudinal PSA data is modeled as a non-linear hierarchical mixed model, with different models for the cured and susceptible groups. To accommodate the heavy tail manifested by the data and possible outliers, we use a t-distribution for the measurement error. The clinical recurrences are modeled using a time-dependent proportional hazards model for those in the susceptible group where the time dependent covariates include both the current value and the slope of post-treatment PSA profile. Estimates of the parameters in the model are obtained by the Markov chain Monte Carlo (MCMC) technique. Residuals from the longitudinal model are plotted for model checking. The model is used to give individual predictors

for both future PSA values and the predicted probability of recurrence up to four years in the future. These predictors are compared with observed data from a validation data set consisting of further follow-up of the subjects in the study. There is good correspondence between the predictions and the validation data.

Individual Prediction in Prostate Cancer Studies Using a Joint Longitudinal-Survival-Cure Model

Menggang Yu[1], Jeremy M.G. Taylor[1 2], and Howard M. Sandler[2]

Abstract

For monitoring patients treated for prostate cancer, Prostate Specific Antigen (PSA) is measured periodically after they receive treatment. Increases in PSA are suggestive of recurrence of the cancer and are used in making decisions about possible new treatments. The data from studies of such patients typically consist of longitudinal PSA measurements, censored event times and baseline covariates. Methods for the combined analysis of both longitudinal and survival data have been developed in recent years, with the main emphasis being on modeling and estimation. We analyze data from a prostate cancer study in which the patients are treated with radiation therapy using a joint model that has been extended by adding a mixture structure to the survival model component of the model. Here we focus on utilizing the model to make individualized prediction of disease progression for censored and alive patients.

In this model each patient is assumed to be either cured by the treatment or susceptible to clinical recurrence. The cured fraction is modeled as a logistic function of baseline covariates, measured before the end of the radiation therapy period. The longitudinal PSA data is modeled as a non-linear hierarchical mixed model, with different models for the cured and susceptible groups. To accommodate the heavy tail manifested by the data and possible outliers, we use a t-distribution for the measurement error. The clinical recurrences are modeled using a time-dependent proportional hazards model for those in the susceptible group where the time dependent covariates include both the current value and the slope of post-treatment PSA profile. Estimates of the parameters in the model are obtained by the Markov chain Monte Carlo (MCMC) technique. Residuals from the longitudinal model are plotted for model checking. The model is used to give individual predictions of both future PSA values and the predicted probability of recurrence up to four years in the future. These predictions are compared with observed data from a validation data set consisting of further follow-up of the subjects in the study. There is good correspondence between the predictions and the validation data.

KEY WORDS: Prostate cancer data; Joint models; Markov chain Monte Carlo; Longitudinal models; Survival models; Cure models.

1 Introduction

There are many circumstances in which both a repeatedly-measured biomarker outcome and the elapsed time to an event are collected on each individual in a medical study. These biomarkers

[1]Department of Biostatistics, University of Michigan, Ann Arbor, 1420 Washington Heights, Ann Arbor, MI 48109, USA.

[2]Department of Radiation Oncology, University of Michigan, Ann Arbor, 1500 East Medical Center Dr., Ann Arbor, MI 48109, USA.

are frequently important health indicators that represent the progression of a disease. Such data will typically have additional features and complications associated with it, including the presence of treatment group indicators and baseline covariates, measurement error in the biomarkers and right censoring of the event time with the possibility of dependent censoring. Joint models for both the marker process and the survival data have been developed in recent years to analyze such data. Estimation of the parameters can be done through a two-stage approach (e.g. Tsiatis et al. 1995, Bycott and Taylor 1998) and a likelihood based approach (e.g. Faucett and Thomas 1996, Wulfsohn and Tsiatis 1997, Henderson et al. 2000, Wang and Taylor 2001, Xu and Zeger 2001, Pauler and Finkelstein 2002, Law et al. 2002). A review on these two approaches can be found in Yu et al. (2004).

PSA is a well known biomarker for prostate cancer used both for screening and for monitoring response to treatment. It is a routine laboratory assay obtained in a blood sample and thus is easy to acquire. Common treatments for patients with local prostate cancer include radiation therapy and surgery. After treatment, clinical recurrence of disease may occur after a period of time. Clinicians and patients monitor the outcome of the treatment by measuring PSA regularly and slight changes or increased values can be a source of great concern and anxiety. In patients who undergo radiation therapy, a sharp rise in PSA after the initial decline is an indicator of treatment failure, and clinical recurrence (reappearance of tumor, either local recurrence or distant metastasis) is expected to follow, although it could be many years before there is clinical manifestation of the recurrence. If the PSA remains low and stable this is an indication that the tumor is not regrowing in the patient. Thus the longitudinal PSA could be useful for predicting cancer recurrence for patients after radiation therapy. The latest value of PSA and the slope of its increase can be very informative about the progression of disease and the hazard of a clinical recurrence. If the pattern of PSA is suggestive of an increased risk of clinical recurrence, the patients may be put on new therapy based solely on this pattern of PSA, typically hormonal therapy with substantial potential side effects, to slow down progression of the disease. Thus methods that enhance early detection of recurrence and accurate prediction of future disease progression for an individual patient based on the pattern of PSA values can have great utility.

A feature in many cancer applications is the fact that some of the patients may have their

tumor completely killed by the treatment, and so will never experience clinical recurrence. These patients are considered to be "cured". We incorporate this aspect of the study into our joint modeling by using mixture cure models (e.g. Farewell 1982, Kuk and Chen 1992, Taylor 1995).

The longitudinal-survival-cure model we adopt here has been used in Law et al. (2002) and Yu et al. (2004). The previously developed model is extended in a number of ways. Besides including additional baseline covariates, we use a time-dependent proportional hazards model depending on hormonal therapy (HT), the current slope, as well as the current value of PSA. To accommodate heavy tails manifested in the longitudinal data and possible outlier, t distributions are used for the measurement error. This model allows estimation of a number of different aspects, including both how PSA changes over time and how this is influenced by other covariates and how PSA influences the hazard of clinical recurrence. However the main focus of the article is on utilizing the model to make individualized predictions for disease progression. Specifically, we predict future PSA values and cancer recurrence probabilities for censored and alive patients. The performance of the prediction is evaluated using a validation data set obtained through further follow-up for these patients.

The task of using a series of biomarker values is considerably more complicated than using a single value, for early detection of disease or monitoring disease progression. The hope is that by using all the data from the serial observations this will lead to an earlier and more precise prediction of future disease progression. There are a few examples of using serial observations in the statistical literature, all involving fairly complicated models and considerable computation. These include using CA125 for early detection of ovarian cancer (Skates et al. 2001), using PSA for early detection of prostate cancer (Slate and Cronin 1997), and using PSA for detecting disease recurrence in prostate cancer (Pauler and Finkelstein 2002).

The rest of this article is organized as follows. In Section 2, we describe two data sets, an analysis data set and a validation data set. In Section 3, we describe a joint cure model. Section 4 presents the Bayesian estimation schemes for the model. Section 5 lists the results. Section 6 uses the model to make individualized predictions and Section 7 assesses the performance of the prediction by comparison to a validation dataset. Finally, we conclude the article with a discussion section.

2 Data

The joint model is developed and fit to one dataset, called the analysis dataset. Predictions derived from this dataset are compared with observations in a second dataset, called the validation dataset, which consists of further follow-up on the subjects in the analysis dataset.

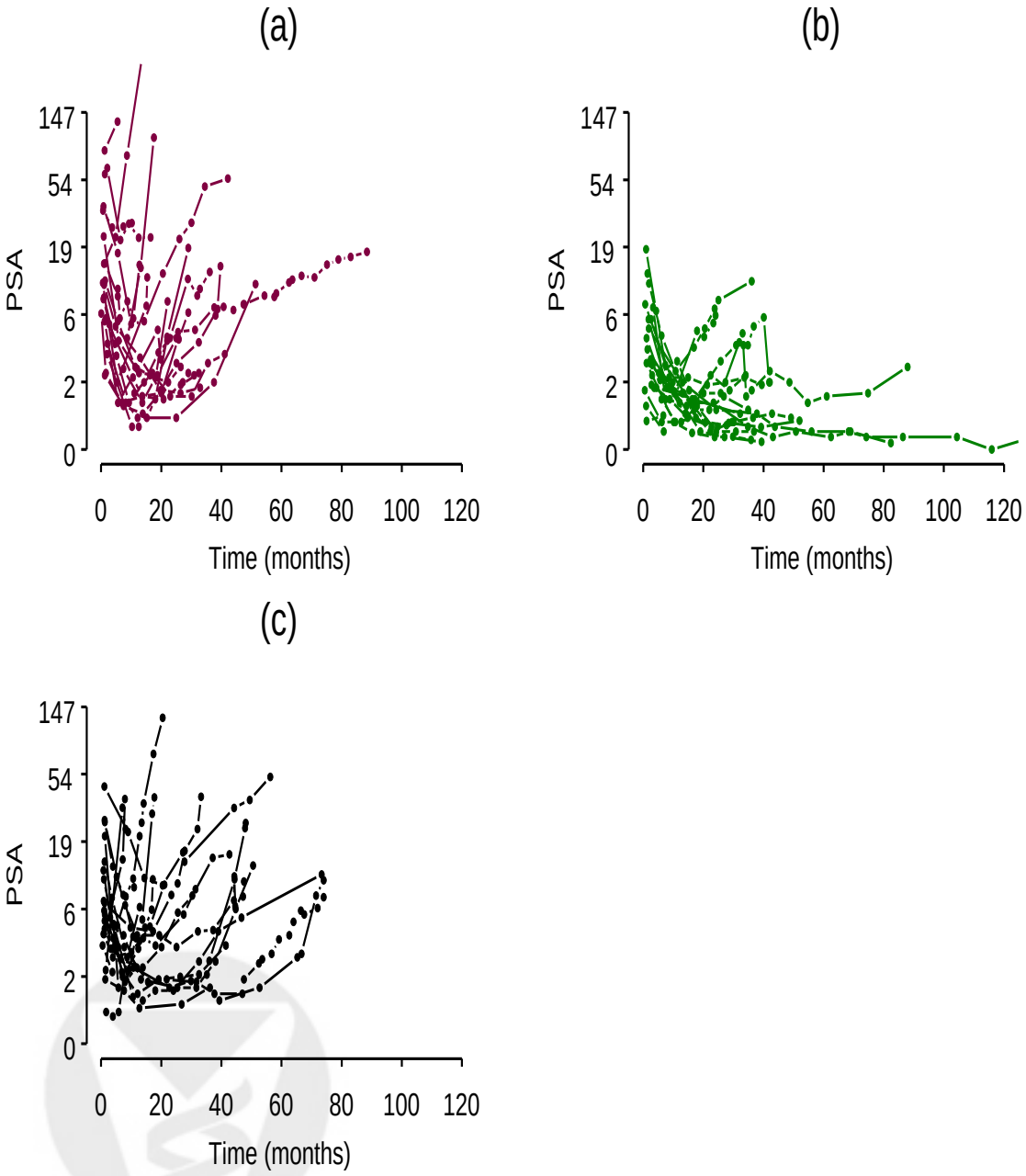
2.1 Analysis dataset.

The data consist of 928 patients with localized prostate cancer, who were treated with external beam radiation therapy at the University of Michigan between July 1987 and February 2000. Patients were excluded from this analysis if they received planned hormonal therapy before the end of the radiation therapy regimen. The baseline variables were age, radiation dose, and duration, T-stage(a measure of the size and location of the tumor), Gleason score (a measure of the aggressiveness of the tumor) and pre-treatment PSA. Earlier version of these data are described elsewhere (Sandler et al. 2000). Post treatment PSA was measured at approximately 6 month intervals. The median number of PSA values per patient was 6 (range 1-29). The total number of PSA measures is 6150. The maximum time between treatment and a PSA measurement is 145 months. During the follow-up period, up to February 2001, 146 patients experienced clinical recurrences, 70 had local recurrence, 72 had distant metastases and 4 had regional failure as their first clinical recurrence. Of the 782 censored patients, 143 died of other causes before any clinical recurrence and 639 were censored at February 2001 or were lost to follow-up prior to that date. Fifty six patients received hormonal therapy prior to any clinical recurrence, of these 15 had a later clinical recurrence.

For patients who were given hormonal therapy, only PSA measured prior to the hormonal therapy are used. For patients who developed clinical recurrence, the PSA measurements before the endpoint are included and for the other (censored) patients, all PSA measurements are included.

Figure 1 are plots of post-treatment PSA measurements for 20 randomly selected patients from each of three categories of patients; those who experienced a clinical recurrence, those who were censored and did not receive HT, and those who received HT prior to a recurrence. We can see a clear pattern of decline in PSA after therapy, followed potentially at a later time by an

Figure 1: Observed post-treatment PSA measurements



(a) 20 patients who had clinical recurrence
(b) 20 censored patients with no hormonal therapy
(c) 20 hormonal therapy patients

increase. The PSA profiles are different among the three groups. We see a clear trend of PSA increase at the later follow-up time after the initial decline for failed patients while we see much flatter curves for censored patients. For patients who received HT as a salvage therapy, we also see a rising trend of the PSA measurements. This is expected because HT is usually given because of a rising pattern of PSA.

2.2 Validation dataset.

The validation dataset consists of all data collected on these 928 patients after February 2001 and available in September 2003. We restrict attention to the 612 patients who were alive at the last contact time in the analysis dataset, and were not known to have experienced a clinical recurrence or received hormonal therapy prior to February 2001. There were 541 patients where new follow-up information was available. Amongst these 541 patients, 472 were alive at the end of the this new follow-up, 63 died from other causes not related to prostate cancer and 6 died from prostate cancer. The median additional follow-up time is 30 months. There were 329 patients with additional PSA values, these patients provided 992 PSA measurements within three years of the previous last follow-up date. Fifteen of the patients developed clinical recurrence in the new follow-up period, 6 of these 15 had HT before the recurrence, an additional 14 have received HT without any clinical recurrence.

3 Notation and Model Specification

Let $\mathbf{z}_i = \{z_{i1}, \cdots, z_{iq}\}$ be the q fixed baseline covariates for subject i. The n_i post-treatment PSA measurements of an individual are denoted by vector $\tilde{\mathbf{y}}_i = (\tilde{y}_{i1}, \cdots, \tilde{y}_{in_i})$, with the corresponding measurement time vector $\tilde{\mathbf{t}}_i = (t_{i1}, \cdots, t_{in_i})$. We denote log transformed post-treatment PSA measurements by $\mathbf{y}_i = (y_{i1}, \cdots, y_{in_i}) \equiv (\log(\tilde{y}_{i1} + 1), \cdots, \log(\tilde{y}_{in_i} + 1))$. This transformation is used so that residuals better satisfy the assumptions of symmetry and homogeneity of variance.

Let t_i be the observed follow-up time, and δ_i be the corresponding censoring indicator. The cure group indicator is denoted by D_i. For a subject i in the susceptible group, D_i is equal to 1; otherwise, it is equal to 2.

Let $\mathbf{x}_{i,obs} = \{\mathbf{y}_i, t_i, \delta_i\}$ be the observed response data for subject i and $\mathbf{X}_{obs} = \{\mathbf{x}_{i,obs},\ i = 1, \cdots, n\}$. Denote $\mathbf{Z} = \{\mathbf{z}_i,\ i = 1, \cdots, n\}$.

Incidence model

The probability of an individual i to be in the susceptible group is given by the logistic function:

$$P(D_i = 1 \,|\, \mathbf{b},\, \mathbf{z}_i) = \frac{\exp(b_0 + b_1 z_{i1} + \cdots + b_q z_{iq})}{1 + \exp(b_0 + b_1 z_{i1} + \cdots + b_q z_{iq})} \tag{1}$$

Longitudinal model

The post-treatment PSA data are modeled by a hierarchical nonlinear mixed effects model. The response model of PSA is given by

$$Y_{ij} \equiv \log(\widetilde{Y}_{ij} + 1) = \log(PSA^*_{ij} + 1) + \epsilon_{ij}, \quad j = 1, \cdots, m_i \tag{2}$$

where $PSA^*_{ij} \equiv PSA^*(t_{ij})$ is the "true" PSA process at time t_{ij} the expression of which is defined below, ϵ_{ij} is the measurement error at time t_{ij}. We assume that the measurement error terms ϵ_{ij} follow a mean 0 t-distribution with degree of freedom $\nu > 0$, and variance σ_e^2 for all $j = 1, \cdots, n_i;\ i = 1, \cdots, n$. We note that the t-distribution can be written as a scale mixture of normal distributions (Lange et al. 1989), that is, we can introduce a latent variable ζ_{ij} such that

$$\epsilon_{ij} \,|\, \zeta_{ij} \sim N(0, \sigma_e^2/\zeta_{ij}) \text{ with } \zeta_{ij} \sim Gamma\big(\nu/2, \nu/2\big)$$

The "true" PSA marker process is modeled by a nonlinear exponential decay and exponential growth model (Zagars and Pollack 1993):

$$PSA^*_i(t) = r_{i1} e^{-r_{i2} t} + r_{i3} e^{r_{i4} t} \tag{3}$$

r_{i1}, r_{i2}, r_{i3} and r_{i4} are the unobserved random effects for subject i (r_{i1}, r_{i2}, r_{i3} and $r_{i4} > 0$). The term $(r_{i1} + r_{i3})$ is the intercept of the post-treatment PSA profile, r_{i2} is the rate of decline of PSA following treatment, while r_{i4} is the rate of rise following the initial decline.

Depending on the patient's cure status D_i, we use different mixed effect model parameters for the true underlying marker profile. For the random effects of a subject i in the susceptible group, we assume

$$\big[\, \mathbf{R}_i \,|\, D_i = 1, \mathbf{z}_i \,\big] \sim N(\mathbf{Z}_i^{(1)} \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \tag{4}$$

where $\mathbf{R}_i$ denotes the log random effects $(\log r_{i1}, \log r_{i2}, \log r_{i3}, \log r_{i4})$ and $\mathbf{Z}_i^{(1)} \boldsymbol{\mu}_1$ is the mean vectors of the random effects in the susceptible group, $\mathbf{Z}_i^{(1)} = (\mathbf{I}_4 \otimes \mathbf{z}_i^*)^T$ is a Kronecker product between $\mathbf{I}_4$ and $\mathbf{z}_i^*$, a vector of baseline covariates.

For the random effects of a subject i in the cured group, we assume that the rate of rise denoted by r_{i4} is close to zero:

$$\begin{cases} \big[\, \mathbf{R}_{i(-4)} \,|\, D_i = 2, \mathbf{z}_i \,\big] \sim N\big(\mathbf{Z}_i^{(2)}\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2\big) \\ \big[\, R_{i4} \,|\, D_i = 2 \,\big] \sim N\big(-6, \sigma_{44}\big) \end{cases} \tag{5}$$

where $\mathbf{R}_{i(-4)} \equiv (\log r_{i1}, \log r_{i2}, \log r_{i3})$, $R_{i4} \equiv \log r_{i4}$, $\mathbf{Z}_i^{(2)}\boldsymbol{\mu}_2$ is the mean vector of these random effects in the cure group where $\mathbf{Z}_i^{(2)} = (\mathbf{I}_3 \otimes \mathbf{z}_i^*)^T$. The mean -6 for R_{i4} for a cured patient is chosen from the fact that PSA level doubles on average in about 20 years for a healthy male. Hence the covariance matrix $\boldsymbol{\Sigma}_2^*$ of $\mathbf{R}_i$ for the cured group is block-diagonal with two blocks, $\boldsymbol{\Sigma}_2$ and σ_{44}. We assume a normal model for R_{i4} to allow for variations of the growth in the cured group.

Conditional failure time model

Conditional on the unobserved random effects, the relative hazard function of the event time t is given by

$$\lambda\big(t \,|\, D_i = 1,\, \mathbf{R}_i, \mathbf{z}_i\big) = \lambda_0(t|\boldsymbol{\eta}) \exp\big[\, \gamma\, PSA_i(t) + \omega\, g\big(sl_i(t)\big) + \kappa HT_i(t) + \boldsymbol{\beta}'\mathbf{z}_i \,\big] \tag{6}$$

where the baseline hazard function is taken to be Weibull $\lambda_0(t|\boldsymbol{\eta}) = \alpha\lambda t^{\alpha-1}$, $g(\cdot)$ is some continuous function, $sl_i(t) = \partial PSA_i(t)/\partial t$ is the slope of $PSA_i(t) \equiv \log(PSA_i^*(t) + 1)$ at time t, and $HT_i(t)$ is a function of t representing HT effect. Our choice of including the slope of the underlying PSA curve is based on commonly used empirical criteria in characterizing prostate cancer progression such as PSA doubling time and PSA velocity. Due to the fact that only a small number of patients receiving HT and an even smaller number of these patients had events, we take a simple function form for HT, $HT_i(t) = 1 - \frac{t-th_i}{a} I(th_i < t < th_i + a)$ where th_i is the time of the HT if the patient has HT, a is a constant to be determined, and $I(\cdot)$ is the indicator function. Hence $HT_i(t) = 1$ when $t = th_i$ and it decreases linearly to 0 at $t = th_i + a$. The reason for taking a decreasing function for HT is that the effect of HT tends to diminish as time progresses.

We explore the appropriate functional form g for $sl_i(t)$, a in $HT_i(t)$, and the degree of freedom ν in the t-distribution through exploratory analysis. As an initial step, we fit a joint-longitudinal-survival cure model to the analysis data set without taking any transformation of slope, that is,

$g(x) = x$, in the conditional failure time model and normal measurement error for the longitudinal data. Then we use results of this modeling fitting to determine g, a, and ν. To choose a value for a, we regard all patients with clinical events or HT, and patients with estimated probability of $D = 1$ greater than 0.6 as "susceptible" and then we fit a Cox model to this group. Then we vary a from 0 to 120 and then compare the partial log-likelihood of each model fitting, and find that 60 may be a good choice for a. To determine a transformation $g(\cdot)$, we use the Poisson and spline-based method described in Therneau and Grambsch (2000) (pg. 120) to explore the functional form for $g(\cdot)$. We find that a square root transformation maybe appropriate. The residuals from the fitting of the longitudinal model is used to determine ν. The degree of freedom 5 is suggested for ν from plots of residuals quantiles versus quantiles of standard normal and t-distributions with various degrees of freedom.

4 Estimation Using Markov Chain Monte Carlo

4.1 Prior Distributions

We fit the model using a MCMC technique. Data driven vague normal priors are taken for $\mathbf{b}, \gamma, \omega, \kappa, \boldsymbol{\beta}$. Specifically, we treat all censored patients with censoring time > 60 months and last longitudinal PSA < 4 as cured. Then we fit a logistic model to all patients in order to get the mean of the normal prior for $\mathbf{b}$, we set the prior variance of each component of $\mathbf{b}$ as 16, which is approximately 100 times the variance estimate from this simple method. Similarly, we obtain prior means of $\gamma, \omega, \kappa, \boldsymbol{\beta}$ by fitting a Cox proportional hazard model to non-cured patients from the above simplified rule and using the nearest preceding value of PSA as the current value. The prior variances are obtained by inflating approximately 100 times the variance estimate from the simpler method.

Vague conjugate priors are used for other parameters. Independent multivariate normal distributions are used as prior distributions for each row of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. The prior of σ_e^2 has an inverse Gamma distribution with mean 1 and variance 10. The prior for $\boldsymbol{\Sigma}_1$ is from inverse Wishart distribution $Inv-Wishart_{\nu_1^0}(\mathbf{S}_1^0)$ with mean $\mathbf{S}_1^0/(\nu_1^0 - 5)$. We take the mean as the estimated $\hat{\boldsymbol{\Sigma}}_1$ from simplified analysis using the longitudinal data alone and the degree of freedom $\nu_1^0 = 20$, which is close to the dimension of $\boldsymbol{\Sigma}_1$. From the posterior distribution of $\boldsymbol{\Sigma}_1$, the prior specifi-

cation has small impact on the posterior distribution. Similarly the prior for $\mathbf{\Sigma}_2$ is from inverse Wishart distribution $Inv-Wishart_{\nu_2^0}(\mathbf{S}_2^0)$ with the degree of freedom $\nu_2^0 = 19$, which is close to the dimension of $\mathbf{\Sigma}_2$.

An inverse Gamma distribution with mean 0.1 and variance 2 is used as the prior for σ_{44}. The reason we use 0.1 for the mean is that there should be less variation for r_{i4} for the cured group. From the posterior distribution of σ_{44}, we can see that the posterior is dominantly determined by data. For the parameter λ of the baseline hazard, the prior is taken from a gamma distribution with mean 0.01 and variance 100. For the shape parameter α of $\lambda_0(t)$, we assume that it has uniform prior on $[0.5, 2.5]$.

4.2 Posterior Distributions and Implementation Details

The posterior distributions for all the parameters can be obtained from the product of full complete data likelihood and prior distributions. The full complete data likelihood is determined by model components (1), (2), (3), (4), (5), and (6) specified in section 3 given the fully observed data $\mathbf{x}_{obs} = \{\mathbf{z}_i, \mathbf{y}_i, t_i, \delta_i, \quad i = 1, \cdots, n\}$.

$$
\begin{aligned}
L \;=\; & \prod_i \Bigg[\Big\{ \prod_j N\big(y_{ij} \,|\, D_i = 1, \mathbf{R}_i, \sigma_e^2/\zeta_{ij}\big)^{I(D_i=1)} N\big(y_{ij} \,|\, D_i = 2, \mathbf{R}_i, \sigma_e^2/\zeta_{ij}\big)^{I(D_i=2)} \Big\} \prod_j \zeta_{ij} \\
& \qquad \cdot f\big(t \,|\, \mathbf{R}_i, \mathbf{z}_i, \boldsymbol{\beta}, \gamma, \omega, \kappa, \boldsymbol{\eta}\big)^{I(\delta_i=1)} S\big(t \,|\, \mathbf{R}_i, \mathbf{z}_i, \boldsymbol{\beta}, \gamma, \omega, \kappa, \boldsymbol{\eta}\big)^{I(D_i=1)I(\delta_i=0)} \Bigg] \\
& \cdot \prod_i \Big\{ h(\mathbf{R}_i \,|\, D_i = 1, \boldsymbol{\mu}_1, \mathbf{z}_i)^{I(D_i=1)} h(\mathbf{R}_i \,|\, D_i = 2, \boldsymbol{\mu}_2, \mathbf{z}_i)^{I(D_i=2)} \Big\} \\
& \cdot \prod_i \Big\{ P(D_i = 1 \,|\, \mathbf{b},\, \mathbf{z}_i)^{I(D_i=1)} P(D_i = 2 \,|\, \mathbf{b},\, \mathbf{z}_i)^{I(D_i=2)} \Big\} \qquad (7)
\end{aligned}
$$

where $N\big(y_{ij} \,|\, D_i = 1, \mathbf{R}_i, \sigma_e^2/\zeta_{ij}\big)$ and $N\big(y_{ij} \,|\, D_i = 2, \mathbf{R}_i, \sigma_e^2/\zeta_{ij}\big)$ are the normal density for transformed longitudinal data from (2) conditioning on their incidence group, $f(t \,|\, \cdot) = \lambda(t \,|\, \cdot) S(t \,|\, \cdot)$ is the density function of the conditional failure time model for subjects in the susceptible group. $h(\mathbf{R}_i \,|\, D_i = 1, \boldsymbol{\mu}_1, \mathbf{z}_i)$ and $h(\mathbf{R}_i \,|\, D_i = 2, \boldsymbol{\mu}_2, \mathbf{z}_i)$ are densities for random effects conditioning on their incidence group from (4) and (5). The expressions for all conditional posterior distributions can be found in Yu (2004).

Adaptive Rejection Sampling (Gilks and Wild 1992) is used for $\mathbf{b}, \gamma, \boldsymbol{\beta}, \omega, \kappa$ since the posteriors are log-concave. A random walk chain (Metropolis *et al.* 1953 and Casella and George 1992)

algorithm is used for random effects $\mathbf{R}_i$. Mean 0 Normal distributions are used for perturbation to get candidate draws. The standard deviations are chosen so that the acceptance rate is about 0.20. Similarly a random walk chain is used for α. The continuous covariates are centered to improve convergence and reduce correlation between them. For censored subjects we drew two sets of values of $\mathbf{R}_i$, one under the assumption that $D_i = 1$ and one under the assumption $D_i = 2$. This facilitated drawing a new value of D_i in the chain.

The program is written in c++. It takes an average of 56 hours to run 15,000 iterations on a Sun workstation. We use both multiple sequences (Gelman and Rubin 1992) method and traceplots for population parameters to check for convergence of the Gibbs sampler in the early exploratory stage of checking convergence. The final results are based on 10,000 draws after 5,000 iterations of burn-in.

5 Results

The main parameter estimates from the model fitting are listed in Table 1. We show the posterior mean and standard deviation derived from the posterior draws of the parameters in the table. We compare the ratio of posterior mean over posterior standard deviation of a parameter and then compare with 1.96 to assess the significance of covariate effects.

Table 1 shows the result for the incidence model (1), we find that a patient's tumor stage, baseline PSA and total dose of radiation are significantly related to the probability of cure in the expected direction. Gleason score, age, and Duration of the treatment are not significant.

For the conditional failure time model, we see that the slope of PSA profile affects the hazard of cancer recurrence greatly. Large slope highly increases the hazard. HT reduces the risk while higher Gleason score and baseline PSA are associated with elevated risk for not cured patients.

We use residuals from the fitting of the longitudinal model, $y_{ij} - \widehat{PSA}_{ij}$, to check the model where

$$\widehat{PSA}_{ij} = \frac{1}{K}\sum_{k=1}^{K} \log\left[e^{R_{i1}^{(k)}} \exp\left\{-e^{R_{i2}^{(k)}} t_{ij}\right\} + e^{R_{i3}^{(k)}} \exp\left\{e^{R_{i4}^{(k)}} t_{ij}\right\} + 1\right]. \tag{8}$$

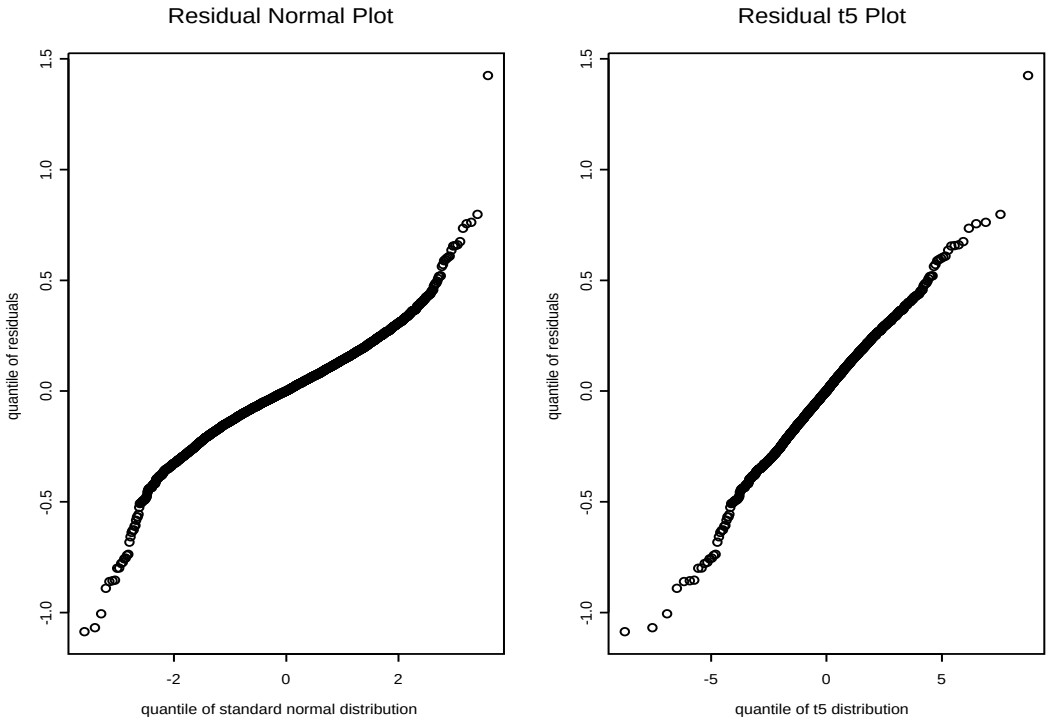
with $R_{i1}^{(k)}, R_{i2}^{(k)}, R_{i3}^{(k)}$, and $R_{i4}^{(k)}$ are from the MCMC output for $k = 1, \cdots, K$.

Table 1: Parameter Estimates in the Joint Models

The incidence model	Est	S.D.	Est/S.D.
Intercept	1.857	0.605	3.070
$I(Tstage = 1)$	-1.512	0.535	-2.826
$I(Tstage = 2)$	-1.518	0.527	-2.881
ln(bPSA+1)	0.769	0.204	3.771
Gleason	0.106	0.094	1.129
Age at RT	-0.011	0.018	-0.641
Total Dose	-0.086	0.039	-2.209
Duration	-0.029	0.022	-1.322
The failure time model			
$I(Tstage = 1)$	-1.133	0.357	-3.172
$I(Tstage = 2)$	-0.337	0.237	-1.423
ln(bPSA+1)	0.230	0.119	1.941
Gleason	0.227	0.077	2.945
Total Dose	0.034	0.031	1.096
$PSA(t)$	0.211	0.086	2.451
Sqrt Slope	5.326	0.769	6.928
HT	-2.865	0.678	-4.228
Baseline hazard			
α	1.093	0.123	8.869
λ	0.002	0.001	1.727
Measurement error			
σ_e^2	0.013	0.005	27.3

We plotted the absolute residuals $|y_{ij} - \widehat{PSA}_{ij}|$ versus predicted log-transformed current PSA from (8), baseline PSA, Gleason score and Tumor stages, we see no clear trend in any of these plots. Figure 2 shows a normal and a t_5 quantile-quantile plots of the residual. We see that the t distribution with 5 degrees of freedom is reasonable. We note that using a t distribution has the ability to accomodate outliers but had little impact on the results of parameter estimates listed in Table 1.

Figure 2: Residual plots for longitudinal model



6 Model Predictions

Suppose we wish to forecast the t_0 months recurrence free probability $P\{\, T_i > t_i + t_0 \,|\, \mathbf{X}_{obs}, \mathbf{Z} \,\}$ for a censored patient i (still alive at the censored time t_i) based on the available data $\mathbf{X}_{obs}, \mathbf{Z}$. Let $\mathbf{\Omega} \equiv \{\mathbf{b}, \sigma_e, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \gamma, \omega, \kappa, \alpha, \lambda\}$ denote the population parameters in the joint models described in Section 3. With K draws $\{\mathbf{\Omega}^{(k)}, D_i^{(k)}, R_i^{(k)} \quad k = 1, \cdots, K\}$ from the posterior distribution $\big[\, \mathbf{\Omega}, D_i, \mathbf{R}_i \,|\, \mathbf{X}_{obs}, \mathbf{Z} \,\big]$, we can approximate $P\{\, T_i > t_i + t_0 \,|\, \mathbf{X}_{obs}, \mathbf{Z} \,\}$ by

$$\begin{aligned} &\frac{1}{K}\sum_{k=1}^{K} P\{\, T_i > t_i + t_0 \,|\, \mathbf{\Omega}^{(k)}, D_i^{(k)}, \mathbf{R}_i^{(k)}, \mathbf{X}_{obs}, \mathbf{Z} \,\} \\ &= \frac{1}{K}\sum_{k=1}^{K}\Big[I(D_i^{(k)} = 2) + I(D_i^{(k)} = 1) P\{\, T_i > t_i + t_0 \,|\, \mathbf{\Omega}^{(k)}, D_i^{(k)} = 1, \mathbf{R}_i^{(k)}, \mathbf{X}_{obs}, \mathbf{Z} \,\} \Big] \end{aligned} \tag{9}$$

where $P\{\, T_i > t_i + t_0 \,|\, \mathbf{\Omega}^{(k)}, D_i^{(k)} = 1, \mathbf{R}_i^{(k)}, \mathbf{X}_{obs}, \mathbf{Z} \,\}\Big]$ is calculated from (6).

We consider two conditional survival curves 4 years after the last contact time for patient i, $P\{\, T_i > t_i + t \,|\, \mathbf{X}_{obs}, \mathbf{Z},$ no HT in $[t_i, t_i + t] \,\}$ and $P\{\, T_i > t_i + t \,|\, \mathbf{X}_{obs}, \mathbf{Z},$ HT at time $t_i \,\}$ for

Figure 3: Individual prediction of distribution of time to clinical recurrence (up to 4 years) for 4 selected censored patients. The left-side vertical axis is shown on a log(PSA+1) transformed scale. The vertical line indicates the time of last contact. The right-side vertical axis shows the probability of being recurrence free from the date of last contact.

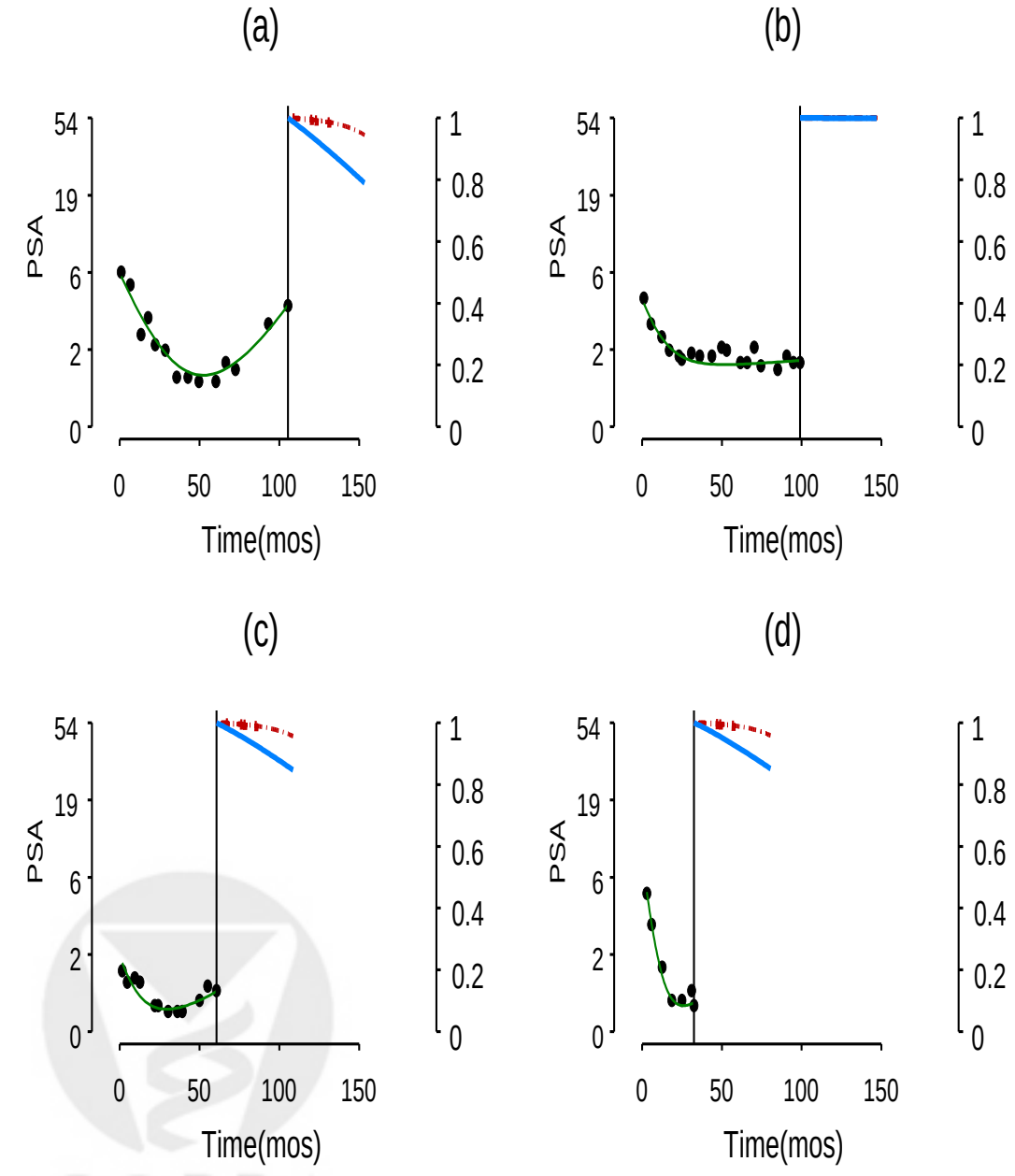
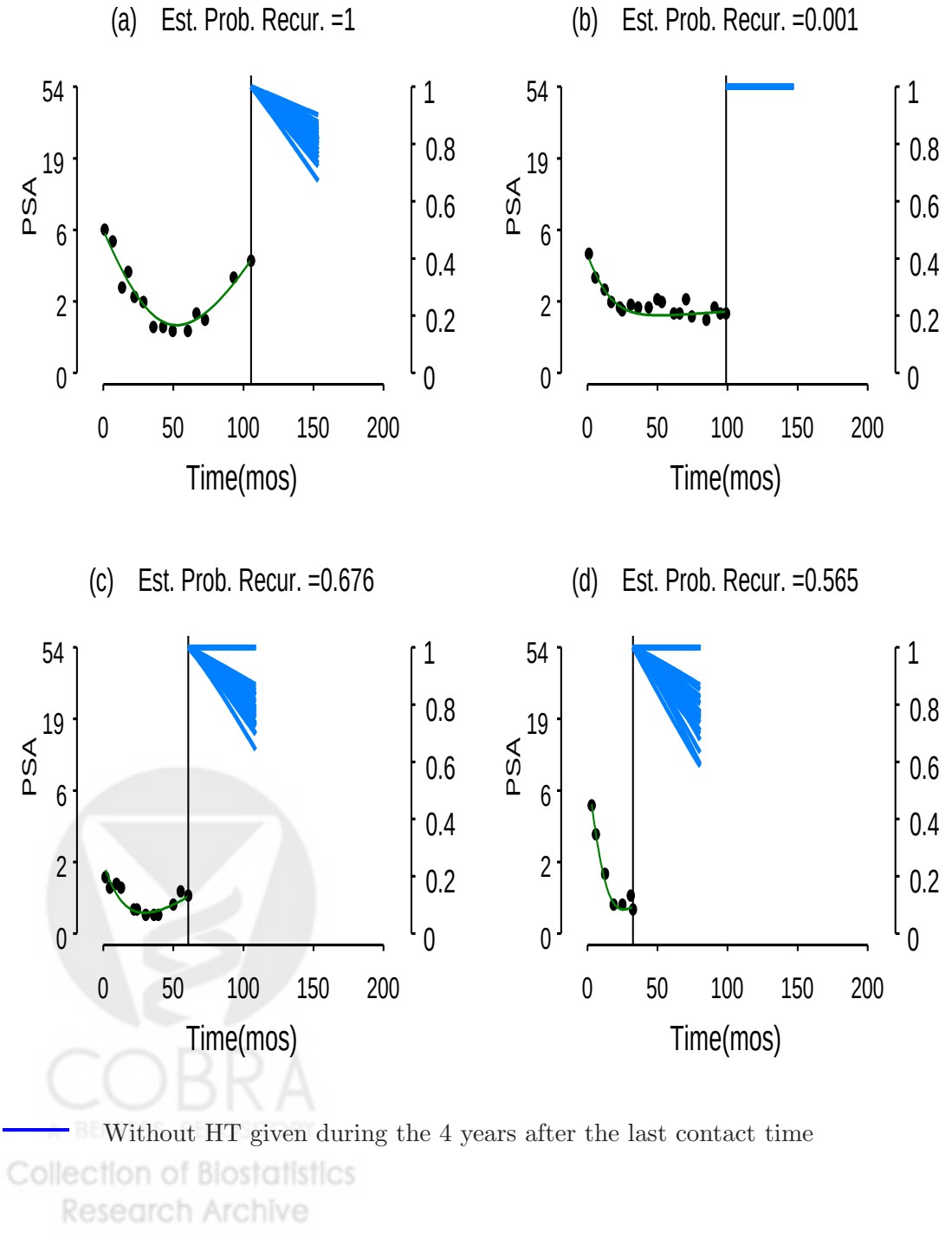


Figure 4: Uncertainty involved with the prediction of distribution of time to clinical recurrence (up to 4 years) for 4 selected censored patients. The left-side vertical axis is shown on a log(PSA+1) transformed scale. The vertical line indicates the time of last contact. The right-side vertical axis shows the probability of being recurrence free from the date of last contact.



$t \in (0, 48]$. By putting these two prediction curves together, we can see the effect of HT on the patient's survival and help aid the decision of whether to give HT for him. Uncertainty of the prediction probability, say, $P\{\, T_i > t_i + t \,|\, \mathbf{X}_{obs}, \mathbf{Z},$ no HT in $[t_i, t_i + t]\,\}$ can be shown by evaluating $P\{\, T_i > t_i + t \,|\, \mathbf{\Omega}^{(k)}, D_i^{(k)}, \mathbf{R}_i^{(k)}, \mathbf{x}_{obs}, \mathbf{z}_i \,\}$ for different draws $\{\mathbf{\Omega}^{(k)}, D_i^{(k)}, R_i^{(k)}\}$ and calculating the variability of the predicted probability.

Figure 3 shows the pattern of PSA and the predicted probability of future clinical recurrence from the date of last contact, with and without the addition of hormonal therapy at the last contact time, for four selected patients. Figure 4 shows the uncertainty of the prediction of clinical recurrence without HT, for the same set of four patients. The estimated probability of eventual recurrence for each patient is also listed in this figure.

Patient (a) and (b) have long follow up time, (c) has medium follow-up time and (d) has relatively short follow up time. The patients were selected to illustrate a range of PSA patterns and predictions. The magnitude of the potential impact of hormonal therapy (HT) can be seen. For patient (a), there is a clear pattern of increasing PSA, suggesting eventual clinical recurrence. Patient (a) has a steep rise which leads to a high probability of recurrence within 4 years. For patient (b) the favorable pattern of PSA post-treatment suggests cure, which corresponds to the almost horizontal predicted clinical recurrence curve. There is little probability of eventual recurrence (0.001) for this patient. Although the PSA values for patient (c) are relatively low, he has a clearly rising pattern, this leads to a quite high probability of eventual recurrence (0.676). Patient (d) has very short follow-up leading to considerable range of predicted probabilities of recurrence within 4 years.

The prediction of future PSA values for patients who are censored, alive and did not receive HT can be calculated from the draws in the Markov chain. For a posterior draw $\mathbf{R}_i^{(k)}$, the predicted (log-transformed) PSA at time t is $PSA_i^{(k)}(t) = \log(r_{i1}^{(k)} e^{-r_{i2}^{(k)} t} + r_{i3}^{(k)} e^{r_{i4}^{(k)} t} + 1)$. By adding corresponding measurement error $\epsilon_i^{(k)} \sim t_5\{0, (\sigma_e^2)^{(k)}\}$, a 95% point-wise predictive interval for log-transformed PSA is then formed using 2.5% quantile to 97.5% quantile of $\{PSA_i^{(k)}(t) + \epsilon_i^{(k)}, k = 1, 2, \cdots, m\}$ for m draws. Examples of these predictive intervals on the original PSA scale are shown as shaded regions for the four patients in Figure 5.

Note that the construction of predictive intervals at each time point is based on the assumption

that HT is not given and that clinical recurrence events and death could be eliminated. This is not the same as the assumption that the patient is alive and is not given HT.

We note that patients (a) and (b), who have lots of data, have fairly narrow prediction intervals, whereas patient (c) and (d) have less follow-up and thus wider prediction intervals. We envision a graph like this would also be useful in monitoring the progression of the patient, for example if a new PSA value is measured and it falls outside the shaded region then this is indicative of the patient doing either worse than or better than expected. After a new measurement is obtained new graphs could be produced, thus giving real-time monitoring of a patient's progression.

7 Validation

7.1 Validation of the Longitudinal Model

The + symbols in the graphs in Figure 5 are PSA measurements obtained from the validation dataset for these 4 patients, all the values fall within the 95% prediction intervals. Patient (c) had a distant metastasis at 61 months after radiation therapy. Patient (a), (b), and (d) have had no clinical recurrence in the validation dataset in the follow-up period. Table 2 shows the proportion of future PSA values amongst all available future data which were within the 95% prediction intervals, we see very good correspondence with the expected 95% level for all years.

Table 2: Prediction and Validation of the Longitudinal Model

Yearly intervals	Total # PSA	Above 97.5%	Between 2.5% and 97.5%	Below 2.5%
0-1 Year	281	1.8%	95.5%	1.1%
1-2 Year	407	3.4%	92.8%	2.9%
2-3 Year	304	3.3%	91.2%	2.3%

7.2 Validation of Conditional Failure Time Model

Table 3 compares the expected events and observed events. For a censored and alive subject i with survival information $(t_i, \delta_i = 0)$ with follow-up information (t_i^*, δ_i^*), the expected number of events within $(t_i, t_i^*]$ is $P\{\, T_i < t_i^* \mid \mathbf{X}_{obs}, \mathbf{Z} \,\}$. Now to calculate the expected number for 1 year, we get the follow-up time for all censored $(\delta_i = 0)$ subjects, $t_i^* - t_i$. If $t_i^* - t_i > 12$, we set $t_0 = 12$, else

Figure 5: Prediction and validation of the longitudinal model

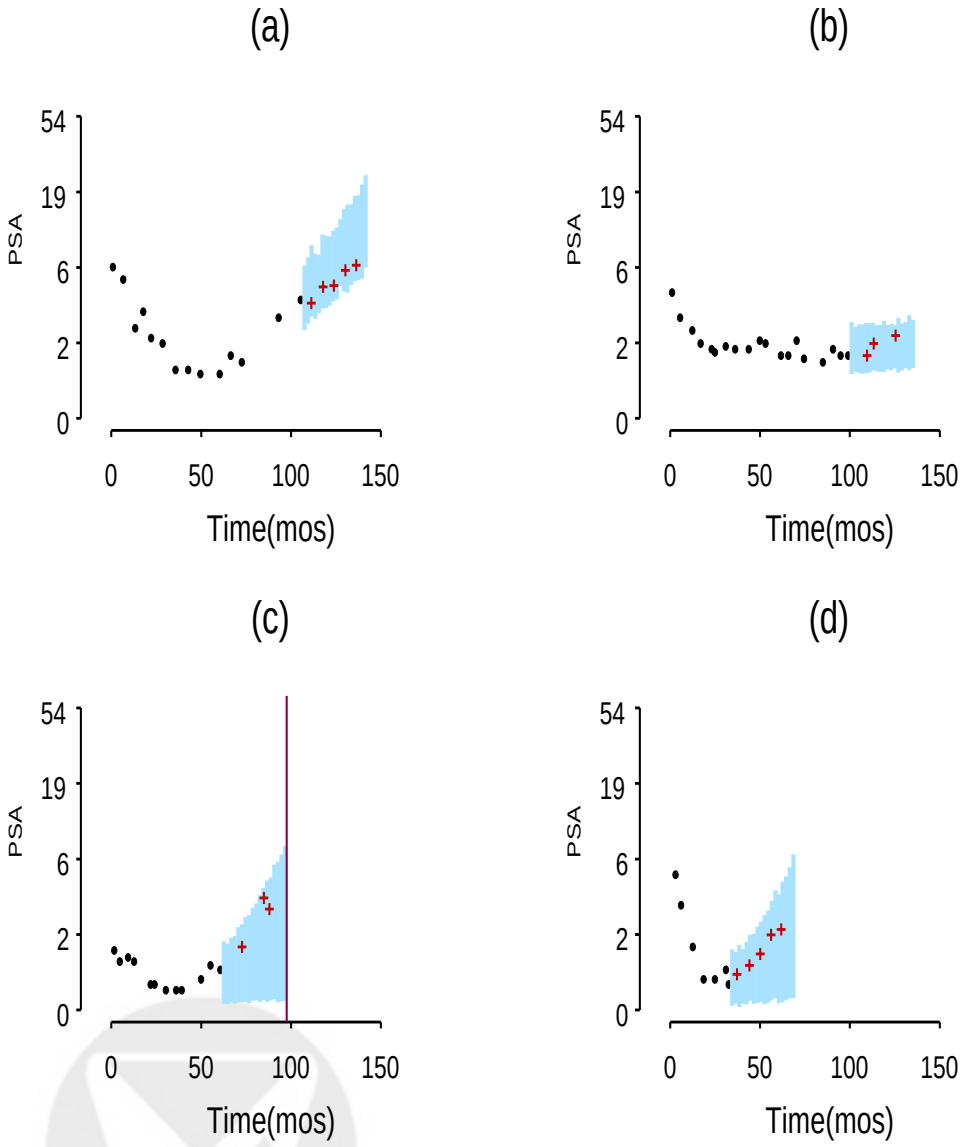


Table 3: Comparison of the Expected Events and Observed Events

	0-1 Year	0-2 Years	0-3 Years
Expected # events	11	20	26
Observed # recurrence	6	10	11
Observed # recurrence or HT	10	20	24

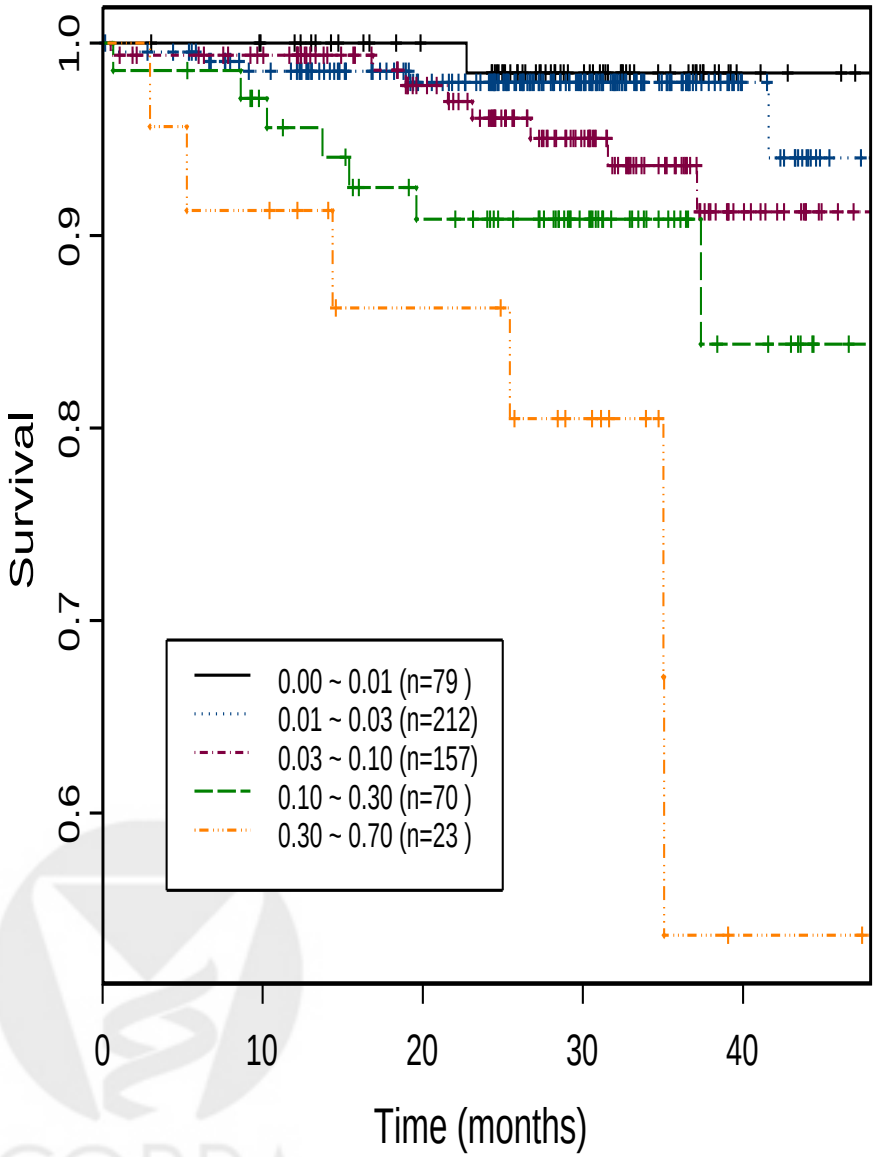
we use $t_0 = t_i^* - t_i$ in $P\{\, T_i > t_i + t_0 \,|\, T_i > t_i \,\}$. By summing over i, we get the expected number of events in $0 \sim 1$ year. Similarly we can calculate expected number of events for $0 \sim 2$ years and $0 \sim 3$ years. From Table 3, we can see that the observed number of recurrence events is much less than the expected number of events for all periods. One reason is that some patients get HT because of elevated PSA. But if we count HT as failure, then the numbers are much closer.

Another way to validate the survival model is to calculate the probability of recurrence within 3 years after the last contact time in the analysis data, $P\{\, T_i \le t_i + 36 \,|\, \mathbf{X}_{obs}, \mathbf{Z},$ no HT in $[t_i, t_i + 36] \,\}$, for any censored and alive patient who had no HT before t_i, and then compare with observed recurrence or HT. The calculated Kaplan-Meier estimate of the three year recurrence or hormonal therapy probability is shown in Figure 6 for five groups categorized by the estimated probability of recurrence within 3 years. The results show that a larger proportion of recurrences or HT patients in the groups with the higher predicted probability, this provides support for the validity of the model. For those who have small predicted probability (< 0.01), there were no recurrence (1 received HT) within 3 years from the last contact time in the analysis data.

7.3 Sensitivity to Priors and Model Assumptions

Due to the large number of patients and longitudinal observations we have in the study, the posterior distributions of population parameters are dominated by data. So we expect our results to be quite robust to most prior specifications. Our limited experience with various prior specifications confirm that this is indeed the case for most parameters. We suspect that the variance term σ_{44} for the distribution of R_{i4} in the cured group, together with the prior mean -6, might have some effects on the classification of the cure status of patients and hence affect the estimates of population parameters. We set a fairly informative prior for σ_{44} to restrict the variation of R_{i4} for the random effects of cured group. Also the mean -6 is chosen rather empirically. These are just

Figure 6: Kaplan-Meier estimate, from the validation dataset of the probabilities free of recurrence or HT. Patients are categorized into five groups by the estimated probability of recurrence within 3 years from the last contact time in the analysis data. Number of patients in each group are listed.



possible ways to separate into states (cured or not cured) for the patients. If we made the prior for σ_{44} less restrictive and the prior mean of R_{i4} larger, we would have more patients with the probability of cure closer to 0.5 since the random effects model under both groups would fit the data equally well. On the other hand, if we set the prior mean of R_{i4} even smaller, we would have more unequivocal estimate of probability of cure, that is, many patients would have estimated probability of cure either closer to 1 or 0. Our experience with various values for both the prior for σ_{44} and prior mean of R_{i4} found very little effect on the population parameters of the incidence model and failure time model. The prior mean but not the prior distribution of σ_{44} affects individual prediction of cure status for some patients. For example, by setting -5 as the prior mean, we have 108 with predicted cure probability <= 0.05 , 36 with predicted cure probability >= 0.95, and 97 with predicted cure probability between 0.45 and 0.55 based on posterior draws of D; while by setting -7 as the prior mean, we have 137 with predicted cure probability <= 0.05, 39 with predicted cure probability >= 0.95, and 84 with predicted cure probability between 0.45 and 0.55. However about 90% of the patients have the difference of predicted cure probability less than 0.1 under the two different prior means. For the prediction distribution of cancer recurrence as described in Section 6, we find prediction curves within three years are nearly always very similar under the assumption of different prior means of R_{i4}.

The advantage of adding a cure model to the joint modeling setting is that it provides a way to model the heterogeneity due to the potential existence of long-term survivors and hence providing more accurate results. To assess the need for a cure model, we advocate using the conditional predictive ordinate (CPO). For a specific subject i under model M_r, the CPO is defined by $CPO_i^{(r)} = f_r(\mathbf{x}_{i,obs} \mid \mathbf{X}_{(i),obs})$, the conditional density of the observed data for subject i, $\mathbf{x}_{i,obs}$, given the observed data $\mathbf{X}_{(i),obs}$ for all subjects except i. The $CPO_i^{(r)}$ can be approximated from MCMC output by a harmonic mean formula (Gelfand 1995),

$$\widehat{CPO_i}^{(r)} = \left\{ \frac{1}{K} \sum_{k=1}^{K} \frac{1}{f_r(\mathbf{x}_{i,obs} \mid \mathbf{\Omega}_r^{(k)})} \right\}^{-1}$$

Computation of this approximation involves evaluation of $f_r(\mathbf{x}_{i,obs} \mid \mathbf{\Omega}_r^{(k)})$ for each draw of the population parameters $\mathbf{\Omega}_r$. For a failed patient, the evaluation of $f_r(\mathbf{x}_{i,obs} \mid \mathbf{\Omega}_r^{(k)})$ requires integration with respect to random effects $\mathbf{R}_i$ and for a censored patient, it requires further integration

with respect to latent variable D_i. Simple simulation studies (without longitudinal data) in Yu (2004) found that CPO has better performance than the Bayesian information criterion (BIC) and the Bayes factors for assessing whether or not a model needed a cure component for survival data. By applying the CPO criteria to our prostate cancer study, we found that the difference in the summation of log-CPO, $\sum_{i=1}^{n} \log(CPO_i^{(r)})$ is 172 for joint models with cure models and without cure models. This supports the need for a cure component in the model.

8 Discussion

In prostate cancer, if the cancer cells are confined to the organ, there is high chance of killing these cancer cells by radiation and hence curing the patient of prostate cancer. These people will not experience recurrence of cancer and the probability of having recurrence is 0. However if cancer cells are not confined to the organ or not completely killed by radiation, then the patient is subject to recurrence. This is the biological reason for including a cure component in the model. Yet we never observe cure for a patient if he is censored (had not experienced recurrence yet). The logistic model provides estimation for the chance of being cured for such patients.

One issue that arises because of patients who received HT due to elevated PSA is dependent censoring. For such patients, had they not received HT, they would very likely experience cancer recurrence soon. The effect of HT postponed the time to recurrence. Adding HT as a time dependent covariate in a hazard model may not be quite correct from a causal inference viewpoint, but it maybe satisfactory for predictions. The decision to give HT is usually based on the value and slope of PSA, and these two variables are already in the model, thus adding HT to the model helps reduce the possible bias since for those patients, the observed time is delayed by HT.

The disadvantage of our model is that it is highly parameterized. With such complicated modeling, interpretation of the parameters can be hard, and furthermore there can be identifiability problems with cure models(Farewell 1986 and Li et al. 2001). On the other hand, the slow progressive nature of prostate cancer also means that recurrences are possible many years after the initial treatment. Thus despite the strong scientific rationale for a cure component, it may be possible to fit these data without using a cure model. However, use of the CPO statistic suggests that models including a cured component give a better fit to the data. In this way the

cure component can simply be regarded as a way to add flexibility to the model.

References

Bycott, P. and Taylor, J. M. G. (1998). A comparison of smoothing techniques for cd4 data measured with error in a time-dependent cox proportional hazards model. *Statistics in Medicine*, 17:2061–2077.

Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, 38:1041–1046.

Farewell, V. T. (1986). Mixture models in survival analysis: are they worth the risk. *The Canadian Journal of Statistics*, 14:257–262.

Faucett, C. L. and Thomas, D. C. (1996). Simultaneously modeling censored survival data and repeatedly measured covariates: a gibbs sampling approach. *Statistics in Medicine*, 15:1663–1685.

Gelfand, A. E. (1995). Model determination using sampling-based methods. In Gilds, W. R., Richardson, S., and Spiegelhalter, D. J., editors, *Markov Chain Monte Carlo in Practice*, pages 145–161. Chapman and Hall, London.

Henderson, R., Diggle, P., , and Dobson, A. (2000). Joint modeling of longitudinal measurements and event time data. *Biostatistics*, 1:465–480.

Kuk, A. Y. C. and Chen, C. H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, 79:531–541.

Lange, K., Little, R. J. A., and Taylor, J. M. G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, 84:881–896.

Law, N. J., Taylor, J. M. G., and Sandler, H. (2002). The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure. *Biostatistics*, 3:547–563.

Li, C. S., Taylor, J. M. G., and Sy, J. P. (2001). Identifiability of cure models. *Statistics and Probability Letters*, 54:389–395.

Pauler, D. K. and Finkelstein, D. M. (2002). Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes. *Statistics in Medicine*, 21:3897–3911.

Sandler, H. M., Dunn, R. L., McLaughlin, P. W., Hayman, J. A., Sullivan, M. A., and Taylor, J. M. G. (2000). Overall survival after psa-detected recurrence following conformal radiation therapy. *International Journal of Radiation Oncology, Biology and Physics*, 48:629–633.

Skates, S. J., Pauler, D. K., and Jacobs, I. J. (2001). Screening based on the risk of cancer calculation from bayesian hierarchical change-point models of longitudinal markers. *Journal of the American Statistical Association*, 96.

Slate, E. H. and Cronin, K. A. (1997). Changepoint modeling of longitudinal psa as a biomarker for prostate cancer. In *Case Studies in Bayesian Statistics III*, pages 435–456. Springer-Verlag, New York.

Taylor, J. M. G. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics*, 51:899–907.

Therneau, M. T. and Grambsch, P. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag New York, Inc., New York.

Tsiatis, A. A., DeGruttola, V., and Wolfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error applications to survival and cd4 counts in patients with aids. *Journal of the American Statistical Association*, 90:27–37.

Wang, Y. and Taylor, J. M. G. (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*, 96:895–905.

Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 53:330–339.

Xu, J. and Zeger, S. L. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Applied Statistics*, 50:375–387.

Yu, M. (2004). *Joint Longitudinal-Survival-Cure Model with Application in Prostate Cancer Studies*. PhD thesis, The University of Michigan, Ann Arbor.

Yu, M., Law, N. J., Taylor, J. M. G., and Sandler, H. M. (2004). Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica*. in press.

Zagars, G. K. and Pollack, A. (1993). The fall and rise of prostate-specific antigen. *Cancer*, 72:832–842.